Glossary of PDF terms

This resource provides end users and non-technical readers with a glossary of the acronyms and terms with lay-person definitions commonly encountered when discussing or describing the Portable Document Format (PDF). Technical readers should always refer directly to the appropriate ISO publication for precise technically accurate definitions. Additional terms are also defined in many PDF ISO standards which can be previewed in the ISO Online Browsing Platform (OBP).

A - B - C - D - E - F - G - H - I - J - K - L - M - N - O - P - Q - R - S - T - U - V - W - X - Y - Z

Action	An action refers to PDF features that enable automatic behaviours triggered by a user interaction or event, such as displaying a diff document when a bookmark is clicked, performing a calculation with form data, or playing a sound or video. For technical details se in ISO 32000-2:2020.
Annotation	Annotations are a special PDF feature most commonly associated with commenting and reviewing a document, such as highlighting strikethrough, or sketching on top of a document. However PDF 2.0 defines 28 different kinds of annotations which provide a far ricl including URLs (Link annotations), watermarks, form widgets, interactive 3D content, redaction, sound, movies and other rich media details see clause 12.5 in ISO 32000-2:2020.
Associated File (or **AF**)	PDF/A-3 (ISO 19005-3:2012) introduced the concept of associated files to PDF 1.7. This "associates" one or more files (typically en PDF object using an **AF** array entry, along with a semantic relationship defined by the **AFRelationship** key in each file specification semantic relationship supports concepts such as *Source*, *Schema*, *FormData*, etc.
AT	AT is most commonly referred to in the context of PDF/UA, and means assistive technology. Assistive technology supports those us to access and navigate PDF documents, such as via screen readers, color and contrast adjustment, screen magnifiers, etc.
Bookmarks	Bookmarks are an informal term for PDF's Document Outline feature. Bookmarks are commonly displayed in a separate navigation document navigation and are a technically distinct feature from headings in content. For technical details see clause 12.3.3 in ISO 3
Conformance Level	PDF Conformance Levels are represented by letter designators with a PDF ISO subset acronym, such as PDF/A-1b, PDF/A-4e, PD 2s. Each Conformance Level relates to a specialized definition in the corresponding PDF ISO subset with its own very specific set c requirements. For example, PDF/A-4 is the PDF-for-archival standard supporting PDF 2.0, with PDF/A-4e being a highly specialized PDF/A-4 supporting engineering workflows with 3D content (hence the "E" designator) while PDF/A-2b (basic) and PDF/A-2u (Unic requirements related to Unicode text extraction capabilities. Not all PDF ISO subsets use conformance levels.
Collection	PDF Collections (or "PDF portable collections") were introduced with PDF 1.7 and are also known by several informal terms such a "package". These special PDF files contain a collection of embedded files in folders and allow the author basic control over the pres collection of files - they can simplistically be thought of as a ZIP container with a user interface.
Conventional PDF	A colloquial term referring to PDF files that do *not* use cross-reference streams or object streams. Such PDFs will therefore always the `xref`, `trailer` and `startxref` keywords.
COS	COS is the acronym for "Carousel Object Syntax", which is the syntax used by PDF and FDF files and is fully described in ISO 320(you see if you look inside a PDF file. "Carousel" was the codename for Acrobat 1.0 when this syntax was first introduced.
Cross-reference table	The common colloquial usage of the term "cross-reference table" refers to information stored in a PDF file below the `xref` keyworc PDF, or in a cross-reference stream (PDF 1.5 and later). However, the formal definition in ISO 32000 defines "cross-reference table" to be a data structure comprising information from all cr sections and cross-reference streams in a PDF file that contains information that permits random access to all indirect objects withir 7.5.4 "Cross-reference table").
Cross-reference entry	A "cross-reference entry" is a formal term of art defined in ISO 32000 for conventional PDFs. It refers to the fixed-length 20-byte line or free objects in each cross-reference subsection.
Cross-reference section	A "cross-reference section" is a formal term of art defined in ISO 32000 for conventional PDFs. It begins with a line containing the keyword `xref` followed by one or more cross-reference subsections. See 7.5.4 "Cross-reference table". The common colloquial usage of "cross-reference table" is often incorrectly used to describe what is technically a "cross-reference s
Cross-reference stream	Cross-reference streams were introduced in PDF 1.5 as a more compact way to define the cross-reference data in PDF. Cross-refe binary data and, because they are standard PDF stream objects, they can use filters and be compressed. Cross-reference streams with object streams.
Cross-reference subsection	One or more "Cross-reference subsections" exist within each cross-reference section of conventional PDF. Each subsection starts containing a pair of integers (defining the first object number and number of objects in the cross-reference subsection respectively), more lines containing the fixed length, 20-byte entries for a contiguous range of object numbers. See 7.5.4 "Cross-reference table"
Direct object	A direct PDF object is an object that occurs inline where it is defined and that does have its own object identifier. In contrast to indire objects cannot be directly referenced as they do not have their own object identifiers.
Fast web view	"Fast web view" is an informal term for the Linearized PDF feature that enables the first page of a PDF file to be available for rapid (rest of the PDF file is fully downloaded (such as while downloading from the internet).
FDF	Forms Data Format is a specialized file format, expressed in the same COS syntax that PDF uses, used for interactive form data th. PDF 1.2. FDF can be used when submitting form data to a server, receiving the response, and incorporating it into the interactive fo used to export form data to stand-alone files that can be stored, transmitted electronically, and imported back into the corresponding form. In addition, beginning in PDF 1.3, FDF can be used to define a container for annotations that are separate from the PDF docu apply. For technical details see clause 12.7.8 of ISO 32000-2:2020.
Form (AcroForm)	PDF supports both interactive and non-interactive forms. For technical details see clause 12.7 in ISO 32000-2:2020. Interactive form in PDF 1.2 as a collection of fields for gathering information interactively from the user and are sometimes referred to as "AcroForm may contain any number of fields appearing on any combination of pages, all of which make up a single, global interactive form spa document. Arbitrary subsets of these fields can be imported or exported from the document as FDF or XFDF. Non-interactive forms 1.7) are a static representation of form fields. Such forms may have originally contained interactive fields such as text fields and rad converted into non-interactive PDF files, they may represent form fields and/or data converted from external sources, or they may h to be printed out and filled in manually.

Fragment Identifier	Annex O in ISO 32000-2 defines PDF-specific fragment identifiers that can be added to the end of URLs that provide anchors to sp influence the display of a linked PDF file. Fragment identifiers are defined by the W3C and appear after the # symbol in a URL. A si that a URL can refer to a specific page in a PDF by appending `page=n` (where *n* starts from 1) to a URL: https://pdfa.org/wp-content/uploads/2019/09/PDF-Association-flyer-A4.pdf#page=2 opens to the 2nd page of this PDF.
Hybrid-reference PDF file	A Hybrid-reference PDF file is a PDF 1.5 (or later) file containing objects referenced by standard cross-reference tables *in addition* streams that are referenced by cross-reference streams. Only PDF 1.5 and later files can be hybrid-reference PDFs because cross-were introduced in PDF 1.5. Refer to clause 7.5.8.4 *Compatibility with applications that do not support compressed reference strea* 2:2020.
Incremental update	A PDF file can be updated incrementally without rewriting the entire file. When updating a PDF file incrementally, changes are appe the file, leaving the original contents unchanged. For example, a PDF-based document review tool may write PDF annotations as in ensuring that a digitally signed original document is not invalidated by the addition of comments. Such technical details are typically users. For technical details see ISO 32000-2:2020.
Indirect object	A PDF indirect object is an object that is defined in the body section of a PDF file with an object identifier (comprising an object num number). It will be referenced elsewhere in the PDF file by using an indirect reference (keyword `R`) with its object identifier.
Integer page index	In PDF the integer page index is a 0-based index of the pages in a PDF file, with the first page having an integer page index of zero used by internal PDF data structures. In contrast, Fragment Identifiers use a 1-based counting system.
Layers	Layers is an informal term for Optional Content Groups (OCGs) in PDF. Layers can typically be individually toggled on and off in int viewers. Examples include architectural drawings where plumbing, electrical wiring, foundations, etc. might each be represented on
Linearized PDF	Linearized PDF is the formally defined feature in PDF feature that enables the first page of a PDF file to be available for rapid displa the PDF is fully downloaded. It is often referred to as "Fast web view". For technical details see Annex F in ISO 32000-2:2020.
Object stream	Object streams were introduced in PDF 1.5 as a more compact method to represent direct objects. Object streams are standard PD (and can use compression filters) and do not contain the keywords `obj` or `endobj`. Object streams are very commonly used with c streams.
OCG	Optional Content Groups are the formally defined feature in PDF which enable selectable layers in interactive PDF viewers. For tec clause 8.11 of ISO 32000-2:2020.
OCR	Optical Character Recognition is the process of recognizing text from an image (photo) of text. It is typically referenced in relation to functionality. The accuracy of OCR results can vary depending on the quality of the page image and other factors. PDF does not co accuracy in any way.
Package	An informal term used to refer to a PDF Collection
Page labels	As documents can be long with many pages, humans have invented conventions to label pages more descriptively to assist with na used to seeing front matter labelled with Roman numerals: i, ii, iii, iv, etc.; appendices prefixed with uppercase letters such as A.1, A chapter/page combinations such as 1-1, 1-2, 2-1, 2-2. In PDF terminology this is what is referred to as a page label - an optional de page that is commonly presented on-screen. This is in contrast to the integer page index used internally in PDF files.
PDF	The Portable Document Format is a random access, binary file format for device-independent, paginated documents that defines ar appearance model for rendering fully typeset text, images and vector graphics. Over time PDF has also expanded to include many specialized features supporting a wide variety of use cases and electronic documents with rich experiences beyond that of "digital p defined by the ISO 32000 family of international standards.
PDF 2.0	PDF 2.0 is the latest version of the PDF specification and is the first PDF specification entirely developed under the ISO consensus formally defined by ISO 32000-2:2020.
PDF/A	PDF/A is an ISO-defined formal subset of PDF designed to support long-term preservation and digital archiving. PDF/A focuses on preservation of the static visual representation of page-based electronic documents over time and is defined by the ISO 19005 fami stands for archival.
PDF/E	PDF/E is an ISO-defined formal subset of PDF 1.6 defined to support the engineering sectors with support for interactive 3D models details see the PDF/E ISO standard ISO 24517-1:2008 *Document management — Engineering document format using PDF — Par (PDF/E-1)*. PDF 2.0 support for engineering workflows is now provided via the PDF/A-4e conformance level - see PDF/A. "E" stand
PDF/R	PDF/R is a small subset of PDF targeting multi-page raster image documents, such as scanned documents. It is based on the PDF Association's PDF/Raster 1.0 specification and is specifically designed to be easy to create in low-end, low-memory embedded dev scanners. It is defined by ISO 23504-1:2020 *Document management applications — Raster image transport and storage — Part 1: (PDF/R-1)*. "R" stands for raster.
PDF/UA	PDF/UA is the ISO-defined formal subset of PDF to support universal access, enabling high levels of accessibility for electronic doc by the ISO 14289 family of standards. "UA" stands for universal access.
PDF/VCR	PDF/VCR enables variable data printing applications using PDF template-based variable content substitution whereby a PDF templ pages with variable content substitution fields (placeholders) is delivered ahead of a print production run and may be reused across production runs, and PDF-based variable data substitution content is provided during print production and merged with the PDF ten final form variable content page output. "VCR" stands for variable content replacement. It is defined by ISO 16613-1:2017 *Graphic t Variable content replacement — Part 1: Using PDF/X for variable content replacement (PDF/VCR-1)*.
PDF/VT	PDF/VT is the ISO-defined formal subset of PDF supporting variable data printing and transactional documents, that builds on the c of PDF/X. PDF/VT is defined by the ISO 16612 family of standards. "VT" stands for "Variable Transactional".
PDF/X	PDF/X is defined by the ISO 15930 family of standards which supports the graphic arts and professional printing sectors. The "X" in eXchange, indicating specialized support for the exchange of digital data targeting professionally printed products.
PDF Declarations	PDF Declarations is an industry-defined specification for declaring conformance to 3rd party standards in XMP metadata. A commo declaring conformance to a specific WCAG level in PDF/UA documents.

PDF version	PDF versions are 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7 and 2.0, with each version defined by its own <u>PDF specification docun</u> generally backward- and forward-compatible, enabling modern software to reliably display old PDFs. Every PDF file identifies a vers file header %PDF-$x.y$, but may also update the version via a special key in the Document Catalog dictionary (Version entry) wher update is applied. Later PDF versions define additional features.
Portfolio	An informal term for a PDF <u>Collection</u>
Revision	An informal term used to refer to an <u>Incremental Update</u>.
Rolemap	Rolemaps are a core <u>Tagged PDF</u> concept that allows any structure type to be conceptually mapped between namespaces in a ma all PDF processors to understand the basic intention of structure types. For example, a custom structure type called Foo might be r paragraph in the standard structure namespace, indicating that semantically Foo is "best matched" as a paragraph. Rolemaps thus custom structure types in PDF. Rolemaps and their use are described in clause 14.7 Logical Structure and clause 14.8 Tagged PDF
startxref	startxref is a reserved PDF keyword that occurs just before the %%EOF end-of-file comment marker along with the byte offset (e: integer in ASCII) to the start of the cross-reference section for conventional PDF files. This keyword is not used in PDFs that only us streams.
Tagged PDF	PDF 1.4 introduced "Tagged PDF" to represent the logical reading order (structure) of a document. It defines a set of standard struc attributes that allow page content (text, graphics and images, as well as annotations and form fields) to be extracted and reused for purposes. <u>PDF/UA</u> uses Tagged PDF to ensure electronic documents are fully accessible. For technical details see clause 14.8 of <u>I:</u>
trailer	trailer is a reserved PDF keyword and defines the start of the trailer dictionary for conventional PDF files. The trailer enables a I quickly find certain special objects and data, such as the largest object number in the PDF (Size entry), the Document Catalog (Rc optional encryption dictionary (if the PDF is encrypted, Encrypt entry). It is an essential part of every PDF file. In PDF 1.5 and later, with the use of cross-reference streams, the trailer keyword does not exist and the trailer dictionary entries cross-reference stream dictionary.
Well-Tagged PDF (or WTPDF)	<u>Well-Tagged PDF</u> is an industry-defined specification for PDF 2.0 that is fully aligned with PDF/UA-2. It allows PDF files to be both r accessible across a wide spectrum of possible use cases.
Widgets	A PDF widget is a specialized type of PDF <u>annotation</u> used with interactive forms and represents the <u>GUI widgets</u> through which da
XFA	XFA stands for "<u>XML Forms Architecture</u>" which is a family of proprietary XML specifications supporting both static and dynamic forr format with limited support in PDF processors, XFA was deprecated in PDF 2.0 (<u>ISO 32000-2:2020</u>) but was permitted in PDF 1.5 -
XFDF	XFDF is the XML equivalent of <u>FDF</u>. It is defined by ISO 19444-1:2019 *Document management - XML Forms Data Format — Part 1* 32000-2 (XFDF 3.0).
XMP	XMP stands for the <u>eXtensible Metadata Platform</u> which is an XML-based standard for metadata used in PDF and required by all IS standards. XMP is defined by ISO 16684-1:2019 *Graphic technology — Extensible metadata platform (XMP) — Part 1: Data model,* *core properties.*
xref	xref is a reserved PDF keyword used to identify the start of conventional cross-reference sections. It is also commonly used collo the phrase "cross-reference". PDF 1.5 and later files that only use cross-reference streams do not use this keyword. PDF files that I <u>updates</u> may have multiple instances of this keyword.